eAppendix for "Mediation analysis with multiple versions of the mediator" by T.J. VanderWeele

Effect decomposition and total effects

The estimators used for the natural indirect effect and the natural direct effect, $Q_1 = \sum_m E[Y|A=1,m,c]\{P(m|A=1,c) - P(m|A=0,c)\}$ and $Q_2 = \sum_m \{E[Y|A=1,m,c] - E[Y|A=0,m,c]\}P(m|A=0,c)$, although they are not equal to the true natural direct and indirect effects respectively when there are multiple versions of the mediator, they do still add up to the total effect because

$$\begin{aligned}
Q_1 + Q_2 &= \sum_m E[Y|A=1,m,c]\{P(m|A=1,c) - P(m|A=0,c)\} \\
&\quad + \sum_m \{E[Y|A=1,m,c] - E[Y|A=0,m,c]\}P(m|A=0,c) \\
&= \sum_m E[Y|A=1,m,c]P(m|A=1,c) - \sum_m E[Y|A=0,m,c]P(m|A=0,c) \\
&= E[Y|A=1,c] - E[Y|A=0,c] \\
&= E[Y_1|A=1,c] - E[Y_0|A=0,c] \\
&= E[Y_1|c] - E[Y_0|c]
\end{aligned}$$

where the third equality follows by iterated expectation, the fourth by consistency and the fifth by unconfoundedness of the exposure. Likewise the true natural direct and indirect effects add up to the total effect essentially by definition: $E[Y_{1M_1} - Y_{1M_0}|c] + E[Y_{1M_0} - Y_{0M_0}|c] = E[Y_{1M_1}|c] - E[Y_{0M_0}|c] = E[Y_1|c] - E[Y_0|c]$. The estimators used for the natural indirect effect and the natural direct effect, Q_1 and Q_2, are no longer equal to the true natural indirect and direct effects respectively since Q_2 captures part of the effect that is mediated.

Proof of Result 3

If data were only available on A, M, Y and C but not on version then the estimators used for the controlled direct effect would be consistent for

$$\begin{aligned}
Q_3 &= E[Y|A=1,m,c] - E[Y|A=0,m,c] \\
&= \sum\nolimits_k E[Y|A=1,k,m,c]P(k|A=1,m,c) - \sum\nolimits_k E[Y|A=0,k,m,c]P(k|A=0,m,c) \\
&= \sum\nolimits_k E[Y|A=1,k,c]P(k|A=1,m,c) - \sum\nolimits_k E[Y|A=0,k,c]P(k|A=0,m,c) \\
&= \sum\nolimits_k E[Y_{1k}|c]P(k|A=1,m,c) - \sum\nolimits_k E[Y_{0k}|c]P(k|A=0,m,c) \\
&= \sum\nolimits_k E[Y_{1k} - Y_{0k}|c]P(k|A=1,m,c) \\
&\quad + \sum\nolimits_k E[Y_{0k}|c]\{P(k|A=1,m,c) - P(k|A=0,m,c)\}
\end{aligned}$$

where the second to last equality follows by assumptions (i*) and (ii*) and the final equality follows by adding and subtracting $\sum_k E[Y_{0k}|c]P(k|A=1,m,c)$. From the definition of F_a we then also have that this is equal to:

$$\begin{aligned}
&= \sum\nolimits_k E[Y_{1k} - Y_{0k}|c]P(F_1=k|c) \\
&\quad + \sum\nolimits_k E[Y_{0k}|c]P(F_1=k|c) - \sum\nolimits_k E[Y_{0k}|c]P(F_0=k|c) \\
&= E[Y_{1F_1} - Y_{0F_1}|c] + E[Y_{0F_1} - Y_{0F_0}|c]
\end{aligned}$$

thus establishing Result 3.

Another way to view the controlled direct effect estimand is thus as two parts where the first part is the controlled direct effect for the exposure on the outcome not through version, standardized by the distribution of the versions of the mediator amongst those with $A=1, M=m, C=c$. The second part is a comparison of the counterfactual $E[Y_{0k}|c]$ standardized by the distribution of versions amongst those with $A=1, M=m, C=c$ versus amongst those with $A=0, M=m, C=c$. This is once again picks up an effect of the exposure on the version of the mediator not through the mediator measure M.